DOCUMENT RESUME

ED 461 351 HE 034 632

AUTHOR Ridley, Dennis R.
TITLE Internet Visibility Measured by Search Engine Hits: A Test on Virginia Colleges and Universities.
PUB DATE 2002-00-00
NOTE 7p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Colleges; Higher Education; Internet; *Online Searching; Universities
IDENTIFIERS *Search Engines; *Virginia

ABSTRACT

This study selected 25 private colleges and universities and 14 public colleges and universities in Virginia. Internet search engines were used to record the number of hits they reported for each of the schools, as a measure of the institution's visibility in the Internet world. Findings indicated that there was a high correlation between visibility on the different search engines, and that visibility varied widely between larger and smaller institutions. (Includes tables showing number of "hits" for each college.) (EV)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 461 351

Internet Visibility Measured by Search Engine Hits:

A Test on Virginia Colleges and Universities

By Dennis R. Ridley, Ph. D.

Virginia Wesleyan College

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Ridley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Inquiries concerning this article should be directed to the author at the following address: Office of Institutional Research and Planning, Virginia Wesleyan College, 1584 Wesleyan Drive, Norfolk/Virginia Beach 23502-5599. E-mail: dridley@vwc.edu.

HE034632

BEST COPY AVAILABLE

Introduction

The growth of the Internet during the past few years, in virtually all of the measurable aspects, has been nothing short of phenomenal. Whether one considers the number of users, the number of sites, the world-wide reach, the volume of business conducted, the interconnectedness and interactivity that characterize Web sites, the amount of scholarly and other research that is published and/or conducted through the Internet—by all of these indicators and more, the same conclusion is reached. The importance and yet-to-be discovered potential of the Internet is one of the most remarkable phenomena ever seen on planet Earth.

However, it remains to be seen whether institutions such as colleges and universities, the traditions of which date from many centuries ago, are prepared to understand, harness and use the power of the Internet. By their nature cautious and conservative, these institutions have a difficult time adapting to, understanding, and using a phenomenon that embodies such rapid change.

Within this broad context, this article makes a modest suggestion: we can begin to understand an institution's visibility in the Internet world by examining the frequency with which the institution is cited within that world. The means of measuring such visibility is readily at hand: Internet search engines, which not only identify but count instances of word strings which correspond to the names of colleges and universities. Thus, for the purposes of this article, "visibility" is defined as the number of hits generated through inputting the college's name into search engines, while controlling for false hits as much as possible.

Test Sample

This study selected 25 private colleges and universities and 14 public colleges and universities in Virginia. All of the institutions selected, while diverse, had two elements in common: they offer four-year undergraduate programs of study or above and are accredited by the Southern Association of Colleges and Schools, Commission on Colleges.

Method

Since the Internet is dynamic and ever growing, it seemed important to capture the data at one moment in time. That moment was on February 10, 2002. Two search engines were used in order to provide an independent measure of visibility as defined in this article: Google.com and Altavista.com.

It was decided that the input would consist of the formal name of the college or university. One exception was made: because of its currency of use, the name Virginia

Tech was used in addition to the more complete and formal, but less used name Virginia Polytechnic Institute and State University.

Both search engines provide the ability to select the language of the sites searched. For both, the option "All Languages" was selected. In addition, there are various options available, which may be studied by anyone who is interested by simply logging onto these web sites and trying several examples. In general, the most literal option was used; that is, the exact string of words found in the inputted name or names was used to match the searched web sites.

Care was taken to avoid ambiguity in order to control for false hits. Such a case was, for example, Bridgewater College, which could bring up "Bridgewater State College." Another example was Virginia Wesleyan College, which could bring up "West Virginia Wesleyan College." There were several other colleges and universities offering the possibility of ambiguity and inflated numbers.

Because of the different procedures required by the two search engines, these situations were dealt with differently for each search engine. Google offered the possibility within Advanced Search to exclude certain key words; for example, by excluding the word "West" from the Virginia Wesleyan College search, all of the West Virginia Wesleyan College entries were eliminated. Altavista offered the same possibility through the use of Boolean logic commands. However, it was often easier and perhaps just as accurate to define the valid entries for that college as those resulting from the following steps: 1) inputting the broader string ("Virginia Wesleyan College"); 2) inputting the narrower string ("West Virginia Wesleyan College"); 3) taking the difference between these two as representing the most valid measure available for the target college. This procedure was followed for Altavista. A consistent methodology was applied so that all similar cases, within one search engine procedure, were treated the same.

The data recorded were simply the number of hits reported by each search engine. For Google.com, this number is given as a round number or approximation, e.g., "about 62,700." For Altavista.com, the number was given as a precise number, not rounded.

Results

The basic results are given below in two tables. Table 1 addresses the number of hits for each of the 25 private institutions. Table 2 shows the number of hits for each of the 14 public institutions.

Each table contains two sorts in descending order of the number of hits. The first has the descending order by Google.com numbers, while the second organizes the results in descending order by Altavista.com numbers.

In addition, three standard Pearson correlations were calculated. The first correlation was between the number of hits for each method, across all private institutions: .86. The

second correlation was across all public institutions: .87. Finally, the correlation across all institutions was calculated: .87.

Discussion

The data may be discussed at different levels. From the standpoint of validity, the high correlations reported above provide reassurance that the two different search engines are measuring the same phenomenon, called here "visibility." Approximately three-fourths of the variance (i.e., the square of the correlations) can be accounted for by the two measures. Such high correlations (.86-.87) are worthy of note.

The high consistency between methods is also shown by the relative absence of anomalous results in terms of inconsistencies. One apparent anomaly among the private institutions was Mary Baldwin College, which ranked fifth by one method but only fifteenth by the other. Similarly, among the public institutions Mary Washington College ranked fourth by one method but only tenth by the other. Such discrepancies might have been mitigated by a little more research and effort at teasing out the valid hits. However, the remarkable fact is that there were so few discrepancies in this exploratory research.

Another observation is the huge span separating the lowest and highest numbers reported, i.e., approximately a factor of 100. The highest numbers, not surprisingly, were reported for large (predominantly public) research universities with extensive research and graduate education resources; the lowest numbers were for small private colleges primarily focused on undergraduate instruction. On the other hand, small colleges and large research universities probably do not need to be compared to each other for most purposes. A more interesting and useful comparison for each might be with their most likely sources of competition, i.e., similar institutions.

The dynamic and ever growing nature of these measures is such that it would be questionable to gauge a single institution's "growth" in Internet visibility by simply taking the difference in measures found at two different times. Indeed, while conducting this study, the author discovered growth of several hundred hits, for the same institution using the same search engine, over the course of several hours during the same day! "Growth" thus measured is more properly interpreted as the phenomenal growth of the Internet itself.

However, it might be meaningful for institutions to compare their ranks and changes in ranks over time, rather than comparing the absolute number of hits. Such comparisons might yield data showing something like a "market share" measure, based on Internet visibility, for similar institutions competing for exposure and attention among those who are paying attention, such as prospective students and their parents. It seems obvious, given the growth of the Internet, that such comparisons can be ignored only at the peril of the institution that is doing the ignoring.

TABLE 1

Virginia Private Colleges: Internet Visibility Measured by Search Engine Hits

Institution	Descending (Google) Search Engines Google	Altavista	Institution	Descending (Altavista) Search Engines Google	Altavista
University of Richmond	62700	292107	University of Richmond	62700	292107
Hampton University	24400	190583	Washington and Lee University	23700	259838
Washington and Lee University	23700	259838	Hampton University	24400	190583
Roanoke College	18500	69583	Roanoke College	18500	69583
Sweet Briar College	16500	22358	Mary Baldwin College	7830	52761
Lynchburg College	15000	33567	Lynchburg College	15000	33567
Randolph-Macon College	13600	13763	Shenandoah University	12000	28502
Liberty University	13200	11170	Sweet Briar College	16500	22358
Shenandoah University	12000	28502	Emory and Henry College	9970	17585
Eastern Mennonite University	11100	10973	Ferrum College	8210	14701
Hampden-Sydney College	10300	7863	Bridgewater College	5840	14463
Emory and Henry College	9970	17585	Randolph-Macon College	13600	13763
Virginia Wesleyan College	8560	10461	Liberty University	13200	11170
Ferrum College	8210	14701	Eastern Mennonite University	11100	10973
Mary Baldwin College	7830	52761	Hollins University	6880	10703
Hollins University	6880	10703	Virginia Wesleyan College	8560	10461
Regent University	6520	6740	Christendom College	3230	10453
Randolph-Macon College Woman's	6500	4390	Hampden-Sydney College	10300	7863
Averett College	6360	5305	Regent University	6520	6740
Bridgewater College	5840	14463	Averett College	6360	5305
Virginia Union University	5720	2777	Marymount University	4720	5221
Marymount University	4720	5221	Randolph-Macon College Woman's	6500	4390
Bluefield College	4510	2203	Virginia Union University	5720	2777
Virginia Intermont College	3240	2326	Virginia Intermont College	3240	2326
Christendom College	3230	10453	Bluefield College	4510	2203

TABLE 2

Virginia Public Colleges: Internet Visibility Measured by Search Engine Hits

Institution	Search Engines (Descending Google)		Institution	Search Engines (Descending Altavista)	
	Google	Altavista		Google	Altavista
University of Virginia	605000	1142253	University of Virginia	605000	1142253
Virginia Tech	532000	492669	Virginia Tech	532000	492669
George Mason University	237000	207982	College of William and Mary	122300	326202
Old Dominion University	135000	73127	Mary Washington College	18800	285164
College of William and Mary	122300	326202	George Mason University	237000	207982
Virginia Commonwealth University	110000	89267	James Madison University	83800	202976
James Madison University	83800	202976	Virginia Commonwealth University	110000	89267
Radford University	44300	71375	Old Dominion University	135000	73127
Virginia Military Institute	26300	67003	Radford University	44300	71375
Mary Washington College	18800	285164	Norfolk State University	13900	69947
Longwood College	18100	26863	Virginia Military Institute	26300	67003
Norfolk State University	13900	69947	Longwood College	18100	26863
Christopher Newport University	13200	20931	Christopher Newport University	13200	20931
Virginia State University	6760	4282	Virginia State University	6760	4282